| RID | Age | Income | Credit Rate | Class Buys Computer. |
|---|---|---|---|---|
| 1 | Youth | Hrgh | Fair | No |
| 2 | Youth | Hrgh | Excellent | No |
| 3 | Middle | Hrgh | Fair | Yes. |
| 4 | Senior | Medium | Fair | Yes |
| 5 | Senior | Low | fair | Yes |
| 6 | Senior | Low | Excellent | No |
| 7 | Middle | Low | Excellent | Yes |
| 8 | Youth | Medium | Fair | No |
| 9 | Youth | Low | Fair | Yes |
| 10 | Senior | Medium | Fair | Yes |
| 11 | Youth | Medium | Excellent | Yes |
| 12 | Middle | Medium | Excellent | Yes |
| 13 | Middle | High | fair | Yes |
| 14 | Senior | Medium | Excellent | No. |

ID3 - Algorithm

Step I : Entropy of class

$$-\frac{P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

P - Probability of success

N - Probability of failure

Step II : Compute expected information requirement for each variable based on the attribute.

$$-\frac{P_i}{P_i+N_i}\log_2\left(\frac{P_i}{P_i+N_i}\right) - \frac{N_i}{P+N_i}\log_2\left(\frac{N_i}{P_i+N_i}\right)$$

Step III : Entropy of attribute.

$$\sum \frac{P_i+N_i}{P+N} \times \text{Information gain of attribute}$$

Step IV : Compute information gain of attribute.

Entropy of class - Entropy of attribute

Step V : Repeat Step 2 for all attributes.

Step VI : Choose the attribute which contains the heighest information gain as a splitting attribute ie, root of tree.

Step VII : Repeat the process until decision tree is formed.

Step1

Entropy of class

From Table; P = Probability of success (Here, No of 'Yes')

= 9

N = Probability of failure (Here, No of 'No')

= 5

Entropy of class buys computes

$$= \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$= \frac{-9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= \frac{-9}{14} \times -0.6374 - \frac{5}{14} \times -1.4854$$

$$= 0.4097 + 0.5305$$

$$= \underline{0.94027}.$$

Step2 : Information requirement of Age.

| Type | Yes | No |
|---|---|---|
| Youth | 2 | 3 |
| Middle | 4 | 0 |
| Senior | 3 | 2 |

i - instance.

Information requirement-

$$-\frac{P_i}{P_i+N_i}\log_2\left(\frac{P_i}{P_i+N_i}\right) - \frac{N_i}{P_i+N_i}\left(\log_2\left(\frac{N_i}{N_i+P_i}\right)\right.$$

Youth

$$= -\frac{2}{5}\log_2\frac{2}{2+3} - \frac{3}{2+3}\log_2\left(\frac{3}{2+3}\right)$$

$$= \frac{-2}{5} \times -1.3219 - \frac{3}{5} \times -0.7369$$

$$= 0.5287 + 0.4421$$

$$= \underline{0.9709}$$

Middle

$P = 4$

$N = 0$

Information gain = 0 (Requirement)

Senior

$P = 3$

$N = 2$

Information requirement = $\underline{0.97094}$.

3. Entropy of Age

$$= \sum \frac{P_i + N_i}{P+N} \times \text{Information requirement}$$

$P_i \rightarrow$ Instance. $P \rightarrow$ Probability of Success

$N_i \rightarrow$ $N$ and Probability of failure.

$$\text{Entropy of Age} = \left[\frac{(P+N)_{youth}}{(P+N)_{class...}} \times \text{Information requirement}_{youth}\right] + \left[\frac{(P+N)_{Middle}}{(P+N)_{class...}} \times \right.$$

$$\left. + \left[\frac{(P+N)_{Senior}}{(P+N)\,class} \times \text{Information requirement}_{Senior}\right]\right.$$

Infor. requir. middle

$$= \frac{5}{14} \times 0.9709 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.9709$$

$$= 0.3467 + 0 + 0.3467$$

$$= \underline{0.6934}.$$

4. Information gain of Age

= Entropy of class - Entropy of ~~Age~~ Attribute.

= 0.94027 - 0.69346

= $\underline{0.24681}$

## Information Requirement of Income.

| Income type | Yes | No |
|---|---|---|
| High | 2 | 2 |
| Medium | 4 | 2 |
| Low | 3 | 1 |

**High**

$$= \frac{-2}{4}\log_2(2/4) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)$$

$$= \frac{-2}{4} \times -1 + \frac{-2}{4} \times -1$$

$$= \frac{2}{4} + \frac{2}{4} = \underline{\underline{1}}$$

← Information requirement

Medium = ~~0.88127~~ 0.9182.

Low = 0.81127

Entropy of Income = $\sum \frac{P_i + N_i}{P+N}$ × Information requirement-

Information = 0.91101.

gain of Income = 0.94027 − 0.91101

= 0.0292

## Information requirement of Credit rate.

| Type | Yes | No |
|---|---|---|
| Fair | 6 | 2 |
| Excellent | 3 | 3 |

Information requirement- ; Fair = 0.81127

Excellent = 1

Entropy of credit rate = 0.89215

Information gain of credit rate = 0.94027 − 0.89215

= 0.04811

| Attribute | Information gain |
|---|---|
| Age | 0.24681 → Attribute containing Highest value. |
| Income | 0.0292 |
| Credit Rate | 0.04811 |

Age

Youth

| Income | Credit Rate | Buys computer |
|---|---|---|
| High | Fair | No |
| High | Excellent | No |
| Medium | Fair | No |
| Low | Fair | Yes |
| Medium | Excellent | Yes |

Middle

| Income | Credit Rate | Buys Computer |
|---|---|---|
| High | Fair | Yes |
| Low | Excellent | Yes |
| Medium | Excellent | Yes |
| High | Fair | Yes |

Senior

| Income | Credit Rate | Buys computer |
|---|---|---|
| Medium | Fair | Yes |
| Low | Fair | Yes |
| Low | Excellent | No |
| Medium | Fair | Yes |
| Medium | Excellent | Yes. |

| Attributes | Entropy | Information gain. |
|---|---|---|
| Age | 0.6984 | 0.24681 |
| Income | 0.91101 | 0.0292 |
| Credit Rate | ~~0.8129~~ 0.89215 | 0.04811 |

→ Attributes having Low entropy and highest information gain is taken as the root node.

→ Entropy increases information gain decreases
↳ degree of randomness increases
We need to split the node again, with respect to another parameter value